TENDENCY OF LAB METHODS USED IN COLORECTAL CANCER RESEARCH DURING LAST 20 YEAR: TEXT-MINING APPROACH

D.Babenko[1]
streampost@gmail.com

[1] *Karaganda State Medical University* (*Karaganda*, *Kazakhstan*)

Key words: Colorectal cancer, Pubmed, text mining, R statistics

Introduction: Colorectal cancer (CRC) is one of the most common cause of cancer death. Investigation of CRC has evolved significantly in the past 20 years. Many lab methods are used to understand mechanism of cancer developing, progression.

A large number of publications about CRC have accumulated now and text mining (TM) has emerged as a potential solution to leverage and exploit large amount of information reported in scientific publication. Here, TM approach was used to extract key words associated with lab technique used in CRC investigation.

Methods: Rentrez r package was used to get PMID number associated with colorectal cancer for 1997-2017.In-house script in the R environment is used to extract title, abstract, date and to create data table. To form pull of unique nouns and adjectives cleanNLP r package with implemented coreNLP library(Standford) were used from which terms associated with lab methods were chosen manually. In the next stage, in-house script were used to extract these terms from each title and abstract.

Results: Pubmed search via rentrez allowed finding 147253 publications associated with colorectal cancer for 1997-2017. There were 513 words/terms associated with PCR, 482 - immunologic with using antibody and interleukins, 26 - chromatography, 21 - with next-generation sequencing, 19 - immunohistochemistry, 11 - electrophoresis, 8 - blotting, 5 - hybridization and 3 - for FISH methods.

As a result, 22.4% (32952 out 147253) of article contained terms associated with lab techniques. The peak of sequencing use is observe in 2015-2017yy. For PCR - 214-2016, for immunologic methods - 2014-2016, for immunohistology - 2013-2015, for hybridization - 2000¬ 2001, for chromatography - 2013-2015, for blotting - 2014-2016, for FISH - 2013-2015, for electrophoresis - 213-2014.

Conclusion: These data showed that modern lab techniques are actively and widely used, with the exception of hybridization. This method has been replaced with sequencing.